AN APPROACH TO EVALUATING THE IMPACT OF A.I.D. PROJECTS

A.I.D. PROGRAM DESIGN AND EVALUATION METHODOLOGY REPORT NO. 5
(Document Order No. PN-AAL-059)

by

Louise G. White
(George Mason University)

U.S. Agency for International Development

March 1986

TABLE OF CONTENTS

PREFACE AND ACKNOWLEDGMENTS

This study describes the methodology characteristic of the impact evaluations conducted by the Agency for International Development (A.I.D.) between 1979 and 1985. The study focuses on the theory and purpose behind this type of evaluation, the procedures that have been developed within the Agency for conducting them, and the ways they have been carried out in practice. It is neither an assessment of the evaluations nor a review of their findings. Nor does this study consider recent evaluations in the impact evaluation approach, such as the broadening of topics to include program and policy reviews and studies of cross-cutting issues or the use of methodologies based on desk studies, brief project reviews, and in-depth fieldwork.

I am very grateful to many within A.I.D. for their assistance and their willingness to give me some of their valuable time. Haven North, Associate Assistant Administrator of the Center for Development Information and Evaluation (CDIE); Marion Warren, formerly Chief, Division of Program and Policy Evaluation (PPE), the unit responsible for the impact evaluations; and Nena Vreeland and Gerald Britan, both of CDIE, all shared some of their preliminary work on this subject and were very helpful in commenting on drafts and discussing the work of their unit. Other CDIE staff who were very gracious in allowing me to interview them at length were David Steinberg, Irving Rosenthal, and Cindy

Clapp-Wincek. Richard Blue, former Director of the Office of Evaluation, was very generous with his time and perceptive in recounting the history of the evaluations. In addition, I interviewed several who had served on impact evaluation teams over the past years: Ross Bigelow, David Dunlop, Steve Lintner, John Wilkinson, Anamaria Viveros-Long, and Pamela Johnson. Their comments served as the source for many of the examples I have included throughout the study and provided me with rich examples of actual field experiences. Finally, I am grateful to Jim Cotter for his careful analysis and critique of an earlier draft and for helping me to clarify the study. While each of these people has made important contributions, the responsibility for particular interpretations is, of course, my own.

Louise G. White
March 1986

1. INTRODUCTION

All evaluations are efforts to learn from experience. Their particular design, however, varies according to their purpose and intended audience. This review describes a particular approach to evaluation that developed within the Agency for International Development (A.I.D.), commonly known as "impact evaluations." Their purpose is to provide program managers with analyses of the results of A.I.D.'s activities and of issues related to the Agency's development assistance policies.

When impact evaluations were formally initiated in l979, the A.I.D. Administrator described them as offering "an in-house capacity to evaluate our work on a regular basis and to produce simple reports which will be of use primarily to us, but also to our host countries, the larger development community, and the Congress." They provide succinct reports of broad, long-range impacts and major development issues, based on field research carried out in a brief timeframe by development professionals.

Impact evaluations have typically included teams of seasoned development experts, some of them A.I.D. staff, who have gone out to the field for 3-4 weeks to make observations, collect information, and write up their analyses. These impact evaluations can therefore more aptly be thought of as "rapid impact studies" to distinguish them from other forms of impact analysis.

This overview describes how this particular approach to impact analysis developed and how it has been practiced during the early 1980s. The discussion can be useful to evaluation administrators, evaluators, and decision-makers. First, it offers a description of the purposes, assumptions, and organization of this approach, which may be of interest to those

who design and administer evaluations. Second, it discusses the validity of the findings, for those who design and conduct the evaluations and also those who use the studies for policy guidance. Third, it provides specific guidance on past impact evaluation methodologies for those who will conduct impact evaluations in the future.

The study emphasizes several characteristics of impact evaluations. First, it is difficult to evaluate long-range impacts of development assistance. Nevertheless, the studies are valuable reminders that despite the problems, it is important to continue examining what difference development assistance has made. In this sense the studies are an important addition to the more usual evaluation of the development process and the means that are used to accomplish purposes. Second, impact evaluations are an important mechanism for helping policymakers throughout A.I.D. address broad policy concerns, thus supplementing other studies by focusing on specific projects. Third, they provide an opportunity to examine not only what has been done, but what else could have been done, and what alternatives might be considered in the future. This stance is less common with more narrowly defined evaluations.

Although directed to policymakers within A.I.D., the studies that have been produced thus far have proved useful to a broader audience. There are several reasons for this interest. First, there is a growing awareness of the value of systematic evaluation throughout the development community. Second, Congress has taken an increased interest in the results of foreign assistance. Third, the reports focus on issues of broad interest that go beyond the results of any specific project. Fourth, those who write them try to write for a general audience and to be as candid as possible. The published studies are regularly distributed to approximately 1,200 individuals and institutions, which include congressional staff, donors, academic and research institutions, private voluntary agencies, developing country organizations, and international bodies.

This report is divided into two parts. The first part, which includes Sections 2 and 3, describes the rationale, activities, and products that typify A.I.D.'s Center for Development Information and Evaluation (CDIE) impact evaluations and that distinguish them from other kinds of evaluations. It is intended for development practioners who want a succinct overview of an impact evaluation process.

The second part, which comprises Sections 4, 5, and 6, goes beyond process to explore more specific methodological and operational issues. It is intended for the narrower audience of those who anticipate direct participation in impact evaluations and who therefore require additional "how-to" guidance.

PART 1.

A DESCRIPTION OF A.I.D.'S IMPACT EVALUATION APPROACH

2. THE NATURE OF RAPID IMPACT EVALUATIONS

Impact evaluations as a specific and separate activity were begun by A.I.D. in the fall of 1979. At that time, the difficulty of determining whether development assistance was making any difference was evident. It was particularly difficult to know whether it was significantly improving the life of the poor. The common response to this difficulty had been to examine the process of carrying out assistance and eschew any effort to determine results. The impact evaluations were established with full awareness of the difficulties involved but in the spirit of "we need to think about results; we can't give up the effort."

Most other evaluations of A.I.D. activities focus on project-specific issues and are carried out on a decentralized basis by USA.I.D. Missions. Many of these project-specific evaluations also examine impacts, but they are usually done during or at the completion of a project, and emphasize more immediate impacts. Rapid impact evaluations have different purposes and characteristics:

They provide information on the long-range effects of development activities, especially the effects on the beneficiary population. They look for any changes that may have occurred and try to determine whether assistance activities made any difference. Thus, they are designed to capture a broader array of factors than are effectiveness studies. The latter studies emphasize how well resources have been transformed into intended results, or in the jargon of evaluation, they compare inputs and outputs.

Impact evaluations are more than a single product. They involve a process that begins well before the actual field research phase and continues once the fieldwork is over. The process includes policy and issue discussions throughout A.I.D., background research, and conferences that reach out to the development community.

Impact evaluations are the responsibility of A.I.D.'s central organization -- the Center for Development Information and Evaluation (CDIE) -- rather than the field Missions.

Impact evaluation teams usually include A.I.D. staff, increasing both the opportunities for the results to influence Agency decisions and for strengthening staff understanding of the development experience.

Finally, impact evaluators demonstrate, particularly to Congress, that A.I.D. is willing to take a critical look at its activities and learn from its failures, as well as to document successes.

Impact evaluations seek to obtain results quickly and to define those results in fairly broad terms. Because this inevitably involves taking shortcuts, two kinds of questions have been raised about the studies. The first is the problem of establishing causality. It is usually difficult to state definitively that a particular assistance activity produced whatever results are discovered. To reach this conclusion, one would have to show that other events did not actually bring them about. This is hard to do when one is looking at long-range impacts, because so many factors could have intervened and contributed to them. The second problem concerns the kinds of data that can be gathered and analyzed quickly. Usually the data are based on retrospective opinions and incomplete observations, and statistical sampling is impossible.

Because of these problems, it is easy to question whether the conclusions of rapid impact studies are valid. The answer lies in the nature of the research that is being carried out. Social scientists traditionally distinguish between exploratory and explanatory research. Exploratory studies are primarily descriptive. They begin with a problem area or intervention and explore its various dimensions. Their purpose is to describe these dimensions and suggest what causal factors are at work and what effects the intervention has. True explanatory research, by contrast, is used in areas in which one can be much more precise about the effects of an intervention and in which it is possible to establish comparisons with similar situations in which no intervention was attempted. The purpose of making this distinction is that impact evaluations are a form of exploratory research. This means that they should not be judged by the extent to which they approximate the controls and rigor of causal or explanatory analysis. They should be judged by how well they accomplish their own purposes -- to describe what has happened, to suggest reasons, to consider alternatives, and to pose interesting questions.

The purpose of impact evaluations then is to explore the terrain, to review what has happened since an intervention or policy was tried, to collect impressions about the extent to which the intervention made a difference, to suggest what problems remain, and to speculate about alternative designs for similar programs or projects. Their function, in other words, is to contribute to more informed decisions, rather than come to a definitive conclusion that project X led to result Y, and that Y would not have occurred in the absence of the project.

It is tempting to respond to this distinction between exploratory and explanatory research in one of several ways. The first is to assume that "anything goes," that exploratory or descriptive research does not have to be conscious of or clear about its methods. This is not the case, and one of the purposes of this report is to suggest ways of improving the design and data collection techniques in this approach to evaluation.

A second temptation is to be overly apologetic and to pepper one's findings with so many disclaimers and caveats that their

thrust is lost. It is important to understand the sources of validity of descriptive studies and what they can and cannot claim.

A third tactic is to try to dress up the results and claim a validity that they do not have. One study, for example, was based on interviews with three people at each of 15 sites, an appropriate approach given its purposes. However, the study then aggregated the interviews and treated them as a single sample group for purposes of statistical analysis. This step was an inappropriate way to treat the data.

Given the specific purposes of exploratory and descriptive studies noted above, what are the grounds for being confident in the findings of impact evaluations? There are four sources of validity:

1. Highly qualified development experts conduct these exploratory studies. The assumption is that if sufficiently knowledgeable people are responsible for them, their validity is enhanced. This point was stressed by the Administrator when he noted that "the best people in A.I.D., even in the absence of sophisticated statistics, can report sound and useful impressions....[T]his exercise will not yield scientifically precise (or even consistent) results, but I believe it will offer us much that is useful." In this sense the validity of the studies depends as much on the experience of development professionals as it does on the data.

2. Conceptual clarity provides a second source of confidence. To what extent does the study define its concepts, examine the assumptions in a project, and consider alternative approaches? It is tempting to evaluate studies purely by how well they marshall concrete data, forgetting that the value of an analysis derives also from its clarification of assumptions and careful definition of concepts. Because impact evaluations include a prefieldwork stage when experts are brought in to discuss issues, there is an opportunity to develop a conceptual framework that can guide the fieldwork. In some instances, the framework can be relatively specific and the evaluation can be used to explore or confirm it. In other cases, the framework needs to be very general and open, and the field research can be used to develop it. Whether done prior to or during the research, conceptualizing the issues surrounding assistance is a central feature of impact evaluations.

3. Comparisons are another reason for having confidence in a study. While impact evaluations are seldom able to establish precise controls, they are usually designed to draw comparisons. For example, they examine clusters of related activities or projects in order to draw comparisons among them. Thus, even when an impact evaluation concerns a single project, that study is part of a group of evaluations dealing with similar projects. When specific conceptual frameworks can be developed, they facilitate such comparisons, and one can be more confident in the results of the study.

4. Finally, confidence in a study is based on the appropriateness and careful design of the data collection techniques. Are they designed to remove bias and to capture a full range of results? Section 5 discusses techniques appropriate to rapid field studies and ways to limit their bias.

Each of these topics will be dealt with more fully in this study. The role of experts will be discussed in Section 3, as part of a description of the organization and procedures used in impact evaluations. Conceptualization and design issues will be discussed in Sections 4 and 6. Data collection techniques will be discussed in Section 5. Where appropriate, examples will be drawn from specific impact evaluations to demonstrate how these four dimensions have been dealt with in practice. The examples are not meant to suggest how evaluations should be done; rather, they are intended to be suggestive by illustrating what teams have done in the field. The overall impact evaluation method is also evolving, and new approaches to incorporating a broader range of comparisons, more in-depth study, and more rigorous economic analysis are being developed and tested. These will be discussed in subsequent reports.

3. ORGANIZATION AND PROCEDURES

This section describes the organization of impact evaluations, who is responsible for them, and who carries them out. It also describes the procedures that constitute the "traditional" impact evaluation process.

3.1 Responsibility

A.I.D. has a highly decentralized evaluation system, reflecting its organizational structure. This means that most evaluations are carried out in the field by USA.I.D. Missions and include monitoring studies, mid-term evaluations, and end-of-project reports. Impact evaluations are the major kind of evaluation conducted by A.I.D./Washington. They are the responsibility of CDIE, part of the Bureau for Program and Policy Coordination (PPC). Because of this organization, the evaluations are carried out by a unit with direct access to policymakers and with little vested interest in the activity they are evaluating.

There are two interesting organizational issues: the relationships that those responsible for impact evaluations have with policymakers and those they have with the A.I.D. field Missions.

First, to what extent are impact evaluations used by policymakers? Rapid impact studies are one vehicle for stepping outside of the boundaries of specific projects, for raising broader policy issues, and for considering alternative approaches

to development assistance. This focus provides a broader view of utilization than is normally the case. Studies can be said to be utilized if their proposals are adopted and if they contribute to a policy debate and help to clarify issues and suggest alternatives. Impact studies probably make their major contribution by doing the latter. It is unlikely that such studies alone can lead to policy changes, but insofar as they inform policy debates, they can have an influence. For this to be successful impact studies need to do two things: they need to relate to the policy concerns currently being debated and considered, and they need to be visible and accessible.

Giving responsibility for evaluations to a central unit makes it easier to connect with the issues uppermost on the minds of policymakers in Washington. In addition, there has been some effort to include policymakers in the original design of the evaluations. This ensures that their questions and concerns are part of the study design, and it also gives them more of a vested interest in the results. Managers of evaluations can also pay attention to congressional testimony by A.I.D. administrators and to other policy statements that indicate where the Agency is headed. In addition to substance, timing is important. Evaluations need to be meshed with the decision-making process so that studies are available when relevant decisions are being made.

Those responsible for the rapid impact studies have been experimenting with various ways to make them more visible and relevant to the policymaking audience within A.I.D.. They are paying more attention to how impact evaluation reports are written, particularly by keeping them short and including abstracts and summaries. They are developing new impact evaluation approaches that focus on strategic and cross-cutting issues as well as more traditional sectoral concerns. They have looked for new ways to communicate the results throughout the Agency, such as including short summaries in regular A.I.D. publications or holding special workshops (see Section 6).

A second issue concerns the relationship between the central evaluation unit and the field Missions. This is increasingly important as more decisions are made in the field rather than in Washington. Because of the expanded role of Missions, those who design the studies are likely to explore new ways to include Missions in the selection of topics for evaluation and to solicit information from them concerning the policy and design problems they are confronting. Some have already tried to involve Mission staff to ensure that the evaluations address issues faced in the field.

The evaluation teams try to maintain a balance between independence of and good relations with the Missions. Missions have to approve the evaluation plan and composition of the teams, and thus need to understand and feel comfortable with the projected studies. It is easy to understand why Missions might view evaluations as fault-finding exercises conducted by central administrators. For this reason, CDIE stresses that the purpose of impact evaluations is not to find fault, but rather to provide

useful assistance to the Missions and to A.I.D. in general. Such reassurances have proved to be critical to the success of the studies. The more Missions perceive that they are being evaluated primarily by their peers and that the emphasis is on learning, the more receptive they are likely to be.

3.2 Evaluation Teams

A second organizational feature of impact evaluations is the inclusion of A.I.D. staff on evaluation teams. To date over 250 AID staff have been involved. To maintain their independence, team members purposely are not chosen from the Mission in charge, and they never have responsibility for the projects or activities they are examining.

The quality of these evaluations often depends on getting busy A.I.D. people to take 4 or more weeks away from their desks. Capable A.I.D. staff have generally been willing to serve on the teams. This is particularly the case when top-level administrators emphasize the importance of the studies and try to meet with the teams. Some administrators within A.I.D. see these impact evaluation teams as a useful training ground for their staff, and one indicated that he tries to arrange for anyone working on his staff to participate on one of the impact teams.

Impact evaluation teams have typically varied from three to six members, have been multidisciplinary, and have included people with expertise in the relevant issues who know the setting. A team with a rural sociologist who knows the area will produce a different report from one composed of economists and engineers, none of whom has ever been to the country in question. Therefore, the composition of teams has a large influence on the kinds of issues that are explored. If beneficiary participation is an important issue, then it is useful to include someone knowledgeable about the literature on this topic. If macroeconomic constraints seem important, then this fact would warrant placing an economist on the team. The recent trend has been toward smaller impact evaluation teams, making the choice of members even more critical.

It has proven valuable for teams to include at least one person knowledgeable about the particular society and culture. For example, a team studying irrigation in the Philippines found it very useful to have someone knowledgeable about the local economy, who knew that even though the irrigation associations being studied were successful in mobilizing local farmers, their debt was increasing faster than their income. Similarly a team studying PL 480 Title I in Sri Lanka included a person who knew the culture well. Whereas the others wanted to focus on the role of language, he convinced them that ethnicity was the major factor, a recommendation that proved to be very useful.

Virtually everyone who has participated in an impact evaluation stresses the importance of selecting people who can

work together and draw on each other's expertise. Practically, this means that the group needs to budget some of their scarce time and plan their schedule so they can meet together regularly. If a team intends to split up and send members to different sections of the country for data collection, they will need to set aside some time at the outset and again at the end to meet and share common experiences. The scheduling of the team's time thus becomes an important aspect of their methodology. For this same reason it is preferable, when possible, to travel together to an area, fan out to collect data for a day or so, and then meet before moving on to a second area.

Teams will often hire local experts. Relations with them vary according to their particular assignments and how closely they are integrated into the team. One team hired two local social scientists as full members of the team; others hire local experts only to carry out specific assignments.

3.3 Procedures

Impact evaluations are best described as an iterative analytic process rather than a single product. Steps in the process include the following:

Topic Selection: Discussions to choose salient topics occur at a high level within the agency.

Topic Coordinator: A topic coordinator is selected to design and oversee a group of studies in a subject area.

Background Studies: Discussion papers and special studies are commissioned to provide background on the topic.

Planning Workshops: Meetings of specialists are held to help define crosscutting issues, identify candidate projects, and develop awareness and support among interested staff within A.I.D..

Project Site Selection: Specific project sites are selected where field studies will be conducted.

Team Selection: Teams of professional development experts are assembled.

Team Workshops: Workshops are held to brief teams, draw up scopes of work, and promote team building.

Fieldwork: The team goes to the field for approximately 4 weeks, carries out research, and prepares a draft report.

Conferences: At the conclusion of the field studies in a single topic area, conferences/seminars are held to review and compare the findings.

Publications: A final report on each field study, longer papers on the findings under each topic and for conference proceedings, and short summaries are typically produced.

The rest of this section describes these steps in the impact evaluation process in more detail.

3.3.1 Topic Selection

If rapid impact evaluations are to contribute to policy formulation within A.I.D., it is important to select topics and activities that provide policymakers with the kinds of information they need and want. Topic areas are chosen by and with senior officials in the Agency, and typically topics are chosen to reflect problem areas or areas receiving special emphasis or a significant proportion of A.I.D. resources. Ideas are also gleaned from country development strategy statements, from annual budget submissions, and from congressional presentations. Over time there has been an increasing interest in choosing topics that are likely growth areas. For example, in l983 CDIE began a systematic effort to set a "forward agenda" based on discussions with senior officials throughout A.I.D.. Some of the topics that have been selected are health service delivery; higher agricultural education, research, and extension; technology transfer; and management of development programs. While earlier impact evaluations primarily focused on sector-specific concerns, the agenda has more recently expanded to include a wider range of cross-cutting issues, policy concerns, and program reviews.

3.3.2 Topic Coordinator

Members of the PPC/CDIE staff are assigned as coordinators for each topic. Their role is to assemble a working group within A.I.D. who have an interest or responsibility in a specific area, to define the topic more precisely, and to write an initial scope of work. This first topic definition must be sensitive to major conceptual issues and to concerns of policy staff and senior officials. At the same time, it has to be sufficiently well defined so that it can be carried out in a rather limited timeframe and so that the various field studies will be roughly comparable. Coordinators are the intellectual and organizational resources for each topic and have considerable autonomy. One brought in a consulting firm to assist with each of the studies on his topic to provide some continuity and comparability. Others have chosen to provide the substantive continuity, preparing discussion papers and scopes of work themselves.

3.3.3 Background Studies

The coordinators can commission two kinds of studies at this

time. Discussion papers deal with generic issues; by drawing on both project documents and the academic literature, they become state-of-the-art papers. These papers try to define the major variables, salient issues, and the most interesting problems in each topic. Examples include the discussion papers for the irrigation, agricultural research, and rural electrification impact evaluation topics.

Special studies deal with narrower questions that are either related, or specific, to the topic under study. They may, for example, take the form of a "desk" study of an interesting project experience. For example, CDIE has published a study of technical, social, and administrative issues of rural water projects in Tanzania, as part of a larger impact evaluation effort in the topic of potable water.

3.3.4 Planning Workshops

Topic coordinators arrange small planning workshops drawing on background and discussion papers. These workshops provide an opportunity to identify crosscutting issues, possible candidate projects for individual impact evaluations, and questions of overall methodology. They also provide an opportunity to specialists from other interested A.I.D. offices to participate in the work of the evaluation series. Based on the workshops and background studies, the topic coordinator will often make substantial revisions in the overall evaluation scope of work.

3.3.5 Project Site Selection

While issues are being clarified, projects are selected for specific study, again with the assistance of the working group. Ideally the working group will select projects that will lead to meaningful conclusions. Because impact evaluations are often done in clusters, projects should be chosen to facilitate comparisons. To the extent that the separate development efforts are similar in important respects and are carefully chosen, the evaluations can explore some actual propositions about development and offer cumulative results. Second, the projects ideally will vary according to the different influences that are being explored. For example, if region seems to be an important factor, projects will be selected from different types of regions.

Such purposive selection is admittedly difficult since few projects fit into neat categories. Projects can vary in their design, the country or region, the amount of the investment, whether they are primarily implemented by the private or public sector, the degree to which A.I.D. is committed to them, the support of the host government, and a variety of other factors. Further, projects are often selected for nonmethodological reasons -- Missions are willing to have them; the timing is appropriate; the country is a priority for the Agency.

Some recent impact studies have adopted a different methodology, sending the same team for shorter visits to multiple sites. This approach provides some advantages in ensuring continuity and enhancing comparability, but less depth, unless local background studies are commissioned.

3.3.6 Team Composition

As noted above, teams are selected to include relevant technical and social expertise and at least some in-country experience. The team leader is given considerable responsibility for planning the evaluation and is almost always an A.I.D. staff member. CDIE provides members with appropriate documents to review.

3.3.7 Team Workshops

Coordinators plan workshops for each project impact evaluation. Workshops assemble the team members and A.I.D. staff concerned with the topic and can include country officials, topic specialists, and development professionals outside of A.I.D.. The purpose is to describe the substantive and methodological issues to the team, to list their specific tasks and make plans for carrying out the evaluation, and to provide an opportunity for the team to develop its specific scope of work. The workshops encourage the teams to explore comparable issues. Once refined, the scope of work must be accepted by the topic coordinator, approved by the Director of CDIE, and agreed to by the relevant Missions.

Each team's scope of work identifies major impacts to be investigated. The briefing workshop also permits team members to assess the kinds of data that may already be available and those that need to be collected in the field, and to reflect on the kinds of evidence that they will have to collect.

Increasingly the teams are being asked to consider if there are data that the Missions, or someone in the field, could collect prior to their arrival. These requests need to be forwarded to the Mission. One team found that this approach worked particularly well when a team member knew of specific resources or contacts and could recommend these to the Mission. If teams can be clear at this stage about the kinds of data they need, the Mission may be able to bring together some relevant studies that have already been done.

3.3.8 Fieldwork

Each field evaluation is a uniquely crafted activity.

Traditionally, most teams have remained in the field 3-4 weeks. Teams usually begin by going to the capital city, making contact with the Mission, reviewing their scope of work, and making any changes demanded by logistical problems. Courtesy calls are made on the Embassy and host government officials as deemed necessary in consultation with the USA.I.D. Mission. At this time, team members make necessary contacts with knowledgeable country personnel and arrange for travel, translation, and the logistics of writing the report. During their initial meetings they may need to refine their data collection instruments based on the kinds of data available. If they decide to use questionnaires or interviews, they should pretest these while they are still together. Ideally, this pretesting will be done away from the capital city and timed so that the team can reassemble, reflect on the results, and make any desired changes.

After going out on site visits, the team may return to a central locale to produce a draft report. Although one person (usually the team leader) drafts the basic report, team members provide required data and analysis as well as individual appendixes. Dissenting opinions can be presented in an appendix. It is crucial that the draft of the report be done before the team leaves the field, because they will quickly become consumed by their ongoing responsibilities once they return. (Section 6 covers the requirements and format of the reports.)

3.3.9 Conferences

Once the drafts for the studies related to a single topic are done, the topic coordinator synthesizes the findings and plans a conference at which these are presented. The conference assembles development professionals, representatives of other donor agencies, academic specialists, field staff, and A.I.D. officials to review the findings and reflect on "what we have learned." This can be an opportunity to involve a large group with diverse expertise; for example, the conference on PL 480 Title I brought together approximately 200 people.

3.3.10 Publications

Several kinds of publications come out of this process.

Discussion papers and special studies are published.

Once impact evaluation reports are completed and acceptable to CDIE staff, they are published. To date more than 60 reports have been published.

There is often a final report that incorporates findings from the evaluations, as well as studies on a topic, conference conclusions, and specific policy guidance.

There are various short summaries and brief write-ups for separate distribution or for inclusion in A.I.D. periodicals. Their purpose is to communicate the evaluation results to those who would be most apt to use them for future policy, funding, and program decisions.

These studies are widely circulated in A.I.D., to Congress, other donors, private consultants, host country officials, and the academic community. The studies also provide an important basis for other CDIE evaluation-applications publications, such as the Project Manager's Reference Guides.

PART 2.

A PRACTICUM ON IMPACT EVALUATION METHODOLOGY

4. DESIGNING EVALUATIONS: CONCEPTUALIZING IMPACTS, DESIGNING RESEARCH, AND SELECTING MEASURES

4.1 Developing Conceptual Frameworks

One of the strengths of rapid impact studies is their attention to the conceptualization of issues. Several aspects of the process encourage this. The process that begins with policy discussions about issues and with background papers provides an opportunity for developing a conceptual framework prior to fieldwork to guide the analysis and keep it from being a purely ad hoc activity. Issues and projects are often selected for their relevance to topics or questions in which A.I.D. has an interest. Second, the original review of the literature and brainstorming within the Agency can be useful in pinpointing major conceptual issues and suggesting specific hypotheses. Third, the time constraints force team members to carefully select the most critical aspects of a project or crosscutting issue they will examine. Finally, because some time has usually elapsed since the project was completed, it may be easier to raise questions about the initial assumptions made by the project or program designers.

Although conceptualizing the evaluation topic is important, there are different views of when it should be done -- prior to or during the evaluation. Some stress the value of laying out a conceptual framework prior to the fieldwork. Such a framework would identify the major variables to consider and propose linkages among them, allowing each study to explore the same questions. To the extent that the projects are also chosen with these questions in mind, the separate evaluations can be aggregated and used to draw some general conclusions.{1} These impact evaluations can be based on a fairly precise framework, because concepts are identified beforehand in the discussion papers and workshops. Results linked to a clear conceptual framework also tend to have more weight with policymakers.

Others, however, are reluctant to set constraints on teams that might inhibit them from pursuing interesting and often unique relationships. They tend to be more impressed with the uniqueness and richness of each case and want to take advantage of this. Coombs, for example, refers to the opportunities and constraints that confront any activity and notes the value of "tailor made" studies.{2} Underlying this perspective is the belief that more can often be learned about what is possible through an understanding of the ways in which a particular assistance effort developed over time and accommodated itself to the existing situation.{3} While they would support efforts to conceptualize the evaluation prior to going to the field, they would preserve considerable freedom for the team to pursue the insights and questions that would arise once they are in the field. In this case some general concepts would be developed prior to the fieldwork to ensure some comparability, but the team's task would essentially be to develop a framework.

Clearly there are tradeoffs here. Perhaps the most that can be said is that the precision of the conceptual framework will vary depending on the topic being investigated. To the extent that those who plan the evaluations believe that enough is known and that it is appropriate to develop a specific conceptual framework ahead of time, the results will be more cumulative. Whereas earlier impact evaluations have tended toward individual case studies, recent trends have emphasized the importance of collecting comparable data.

In any case, practical problems in the field situation will often require teams to make adjustments, revising or simplifying initial propositions or choices of variables.

> A team doing an evaluation of irrigation projects had to modify their initial scope of work once they reached the field and got a more realistic feel for the logistics of visiting projects. While still in Washington, they met together for 2 days to review and decide on which of the many possible factors they should focus. After arriving in the country, they decided not to look at changes in income or productivity because there had not been enough time for these to become apparent. They also decided not to look at one of the projects because relatively little had occurred there. And, perhaps most important, they decided not to look at the irrigation systems from an engineering perspective to see if they had been well constructed. They decided to assume that if the projects were providing water this was an indication that they were adequately built.

These omissions allowed them to focus on who had access to the water and to technical assistance, and to deduce from the answers to these questions whether irrigation should have been constructed in this situation at all, and what the results were.

Note the tradeoffs here. To the extent that such changes are made in the design once the team is in the

field, the results of all the studies in a group will be less comparable. Thus, when evaluations do depart from plans developed at the workshop, it is helpful to give reasons why the original questions did not prove useful.

There are three components of a conceptual framework:

Deciding which impacts to examine and describe
Explaining the relationships among the variables
Selecting measures or proxies for the variables

{1}Elinor Ostrom, Strategies of Political Inquiry (Beverly Hills, California: Sage, 1982), pp. 179-222.

{2}American Council for Voluntary Agencies for Foreign Service (ACVAFS), Approaches to Evaluation: Report of a Workshop on Impact Evaluation (New York, New York: ACVAFS, 1981), p. 6.

{3}See Trudi Miller, Public Sector Performance (Baltimore: Johns Hopkins University Press, 1984), and Samuel Paul, Managing Development Programs (Boulder, Colorado: Westview Press, 1982).

4.2 Identifying Impacts

Impacts are commonly differentiated from the outcomes of an activity. They are the long-range results of particular outcomes. For example, in a road building project, the miles of road built would be an outcome, whereas the usage of the road and the effect on living standards and on the economy would be examples of impacts.

Describing impacts is also different from determining if a project was effective. Effectiveness is a measure of whether an activity accomplished its goals, but this can be a limiting question for several reasons. First, some of the most interesting results of a project or activity can be unintended, and therefore studies need to cast a wide net and examine impacts in the broadest sense of what happened as a result of a project. Second, impact evaluations try to raise the issue of whether the project or program should have been designed as it was, and not simply whether it accomplished its purposes. One observer even noted that the term "impact" is limiting, that he likes to use the French word "temoins," meaning a witness to what has happened. Typically, impact studies are done after some time has elapsed, and thus evaluators are able to bring more perspective to their

studies than is usually the case in studies designed primarily around project goals.

Given the emphasis on using impact evaluations to inform policy decisions, concerns within A.I.D. inevitably influence which impacts a team will look for. Since rapid impact studies were first begun, there has been an emphasis on examining the impact of assistance on beneficiaries, particularly on the rural poor. Did improved incomes change the lives of the rural people; did it empower them; did it encourage them to organize to achieve other benefits? Evidence that assistance was easily diverted from helping the poorer members of society was one of the original reasons for developing this style of evaluation, and it has continued to be a major part of the conceptual framework in most studies. For example, one list of suggested impacts included the following:

-- The physical outputs of the project are maintained.

-- The project produces benefits for the poor.

-- The benefits are sustained.

-- The benefits encourage self-sustaining growth.

Determining whether benefits reach the poor is frequently difficult. Often people who seem to be poor by some standards are actually in the top half of the income distribution in a community. One analytical technique is to divide the income distribution in the community approximately in thirds and then determine into which category the beneficiaries of a particular project fall.

A.I.D. policy has recently added other impacts to this list. One is a concern for local institutions and whether assistance increases or impedes their capacity to function. Another is the impact of aid on the private sector. As these different policy interests are incorporated into the conceptual frameworks, they will affect how a study is designed. For example, consider the implications of different ways to define the impacts of a rural electrification project. If the major concern is with the effect on poor people, a study could examine household hookups and try to determine which groups benefited. If there is interest in the impact on the economy, the study could focus on whether the power grid was expanded to market towns. To the extent that A.I.D. is concerned with the effect on private sector initiatives, the same study could look at how the electrical system affects small-scale entrepreneurs.

Some impacts are more difficult to collect information on than others, and often these impacts are the ones of most interest, such as improved health or living standards. In such cases, studies often rely on an examination of outputs -- what was done rather than what difference it made. This approach may still provide useful information. As Judith Tendler puts it:

Potable water is a case where the evaluator will usually not be able to assess the impact of a particular water project on the incidence of disease; but she can make an important contribution by finding out about the percent of the population covered, the quality of the water and, that greatest of problems, the maintenance of the water site and its equipment. Adequate coverage of these areas will often give the information necessary to speculate intelligently about impact.{4}

In considering impacts, it is important to make distinctions among different kinds of impacts rather than saying that "on the average" a project was successful or failed. Impacts can be categorized in the following ways:

1. Immediate and Long-Range Impacts

Two evaluations of a Nigerian training project done at the time of the project's completion focused on immediate results and concluded that the project had failed. Impact studies done after 10 years, however, found that many of the problems cited in the earlier study had been dealt with; from this perspective, the project could be judged a success.

2. Intended and Unintended Impacts

An impact evaluation of a large health program found that unintended impacts were critical in explaining why an apparently successful project was an eventual failure. The project was successful in providing needed health services cost effectively. However, many people in the communities were opposed to the health project. The team decided to expand their analysis to explore the reasons for this opposition. This led them to look at the unintended effects of the project on other health service providers -- traditional healers and private physicians. It became clear that the project affected them adversely, which explained their opposition and why it undermined the project.

3. Positive and Negative Impacts

PL 480 programs, which distribute surplus food to developing nations, are controversial because they can have an adverse effect on domestic food production. One study discovered that people did not want to find such negative impacts and hence did not look for them. The team found it was better to identify a wide range of impacts, both negative and positive, and then determine whether the positive benefits offset any negative impacts.

4. Macro (sector) and Micro (household, firm) Impacts

A study of rural roads looked for the following macro impacts: marketing and subsistence patterns, social services, environmental effects, and general consumption. They also looked at some micro impacts: how farmers used

their land, composition of their crops, and their aspirations.

The point of these examples is to emphasize that identifying impacts and their critical characteristics is a major aspect of the evaluations; time must be set aside for this task both before the team leaves Washington and in the field.

{4}Judith Tendler, Turning Private Voluntary Organizations Into Development Agencies: Questions for Evaluation, Program Evaluation Report No. 12 (Washington, D.C.: Agency for International Development, 1982), p. 149.

4.3 Designing the Study and Developing Working Hypotheses

A second aspect of the conceptual framework is to consider why the impacts occurred. Whether the goals were attained or the impacts were positive or negative, it is useful to develop a conceptual framework that proposes other variables that may be related to the impacts. Because development is a complex process in which myriad socioeconomic, political, institutional, and technological factors interact, it is seldom possible to identify any single cause or even trace a simple linear causal process. This limitation is particularly a factor in impact evaluations undertaken after some time has elapsed, when many other changes in the environment have occurred that make a determination of causality more difficult. The point of noting this complexity is to emphasize that any causal analysis is going to be exploratory and suggestive rather than definitive proof of causality. There will probably always be "other factors" that have intervened.

There are several aspects to this part of the conceptual process: examining internal and external causes, conceptualizing multiple causal chains, developing working hypotheses, and sorting out the influence of the development assistance activity itself.

4.3.1 Examining Internal and External Variables

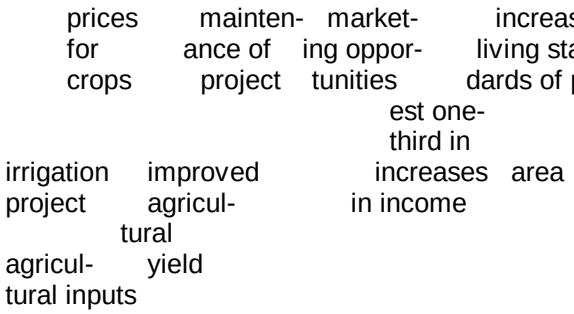
One step in developing a conceptual framework is deciding how much attention to give to factors external to the project as opposed to the characteristics of the project itself. The actions of other donors, price changes, relationships to other assistance efforts, political support, and compatibility with cultural values are all external factors that can affect the impact of an activity. Other kinds of evaluations that are more closely tied to specific projects have tended to focus on the internal characteristics of the project. For example, an evaluation of an irrigation system would determine if it helped farmers grow more rice. Impact evaluations, however, are more

apt to take external conditions into account and include such factors as the prevailing prices for other crops and marketing arrangements.

> A study of manufacturing activities by small entrepreneurs included external factors such as labor market, import controls, credit flows, and mechanisms for spreading and reducing risk, as well as internal factors such as technical knowledge in the firm, degree of specialization, component manufacturing and assembly, and feedback about technology to users.

4.3.2 Identifying Causal Chains

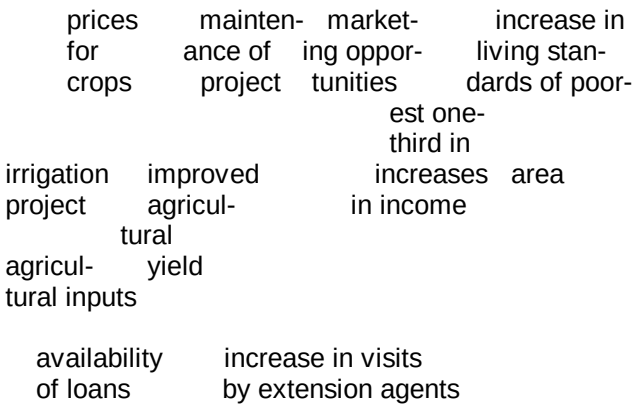
Those planning the evaluation may find it helpful to diagram the presumed relationship among the variables to be examined. For example, consider the following diagram of an irrigation project:



The elements in this diagram could be garnered from the original logical framework of the project, discussion papers, and the professional expertise of those in the working group and evaluation team. Once such a framework is developed, it is possible to decide which factors and relationships to study. Ideally, this framework will help determine which specific sites to include in the study. For example, if "maintenance of projects" is presumed to be important, then it would be useful to select some cases in which maintenance plans did exist and some where they did not, and compare the impacts in the two kinds of situations over time. Or it may not be clear exactly how marketing opportunities affected income, and the team could then compare the impact of the irrigation project in two communities that differ according to their access to an effective market.

Besides helping the team clarify assumptions, this can be a useful technique for exploring alternatives to the initial project design. To be effective, it may be that the irrigation project should have been combined with some marketing innovations or with a program to encourage maintenance. Or the data may indicate a positive relationship among the different elements in the chain but suggest that the poor did not benefit. This might suggest that the project should have included a mechanism for stimulating loans to the poorest or that a more intensive, targeted use of extension agents was needed. Note how these

factors could be added to the first diagram:



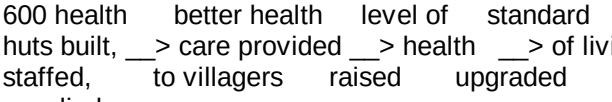
These additional variables may be uncovered during initial discussions in the workshop, or they may only become apparent during the field visits. To the extent that they can be anticipated ahead of time, the various evaluations in each cluster can focus on the same variables and relationships.

It may be possible to develop a general framework ahead of time. For example, the following three sets of variables can be adapted to most projects. The first is the assistance activity itself -- the kind of good or service it is providing. The second set is the institutions that administer or manage the activity. And the third set of variables is the relevant characteristics of the environment or context.{5} Such a framework, if used across several evaluations, would allow A.I.D. to generalize across studies about the patterns that encourage successful activities.

Applying this to the above example, the framework would first describe the nature of the assistance to the small entrepreneurs. Was it offered to individuals? Was it in the form of training or loans? Second, it would examine how the assistance was provided and managed. Was it done through the local government, through the private sector, through a newly organized unit? And third, it would examine the institutional environment, such as the labor market, import controls, and marketing opportunities. Ostrom would go one step further and draw from existing knowledge to propose how the variables are related to each other. This approach would be somewhat more deductive than the example above. Such a schema provides a broad framework and a means for comparing the results of different studies.

The following example illustrates the use of an initial framework that has been revised where necessary.

A study of a rural health care project was based on the following causal chain:



> The team found evidence for the first two items in the chain, but the last two items could not be determined. On closer examination, they also found that many of the health huts were about to close. They suggested that other factors should be looked at, such as the financial management of the projects, supervision by the government, and the continued supply of medicines. Thus, they were able to propose that future health projects be planned with a different causal chain, one that considered the long-term management of the project.

{5}Adapted from Ostrom, pp. 179-218.

4.3.3 Developing Working Hypotheses

In the traditional social science lexicon, specific hypotheses are derived from theory and then tested. Hypotheses can be developed as a guide for rapid impact studies, but they can more accurately be thought of as "working hypotheses" or "propositions" rather than formal causal hypotheses. These terms suggest that the hypotheses reflect operating assumptions rather than a body of tested theory. They also suggest a certain tentativeness and recognition that there are probably other unidentified influences operating. The concept of a "working hypothesis" indicates that the evaluation is one piece in an ongoing learning process and that the hypothesis or proposition will be clarified and improved over time rather than disproved or confirmed. The terms also fit with the view of impact evaluations as contributing to an ongoing exploration of alternative ways to pursue development, rather than a mechanism for determining the effectiveness of project X.

Some hypothesized relationships are fairly predictable, others will be more surprising. It is more useful to spend time explaining the latter. For example, Tendler notes:

> If there is a lack of coordination between government agencies, then little time should be spent on describing the problem because it is so familiar and there are good reasons for it. If coordination is achieved, then a lot of time should be spent explaining how that happened, since it is so unusual. Similarly, if a road does not get maintained, not much time should be spent deploring it. Instead, attention should be paid to the community reaction to the lack of maintenance, and what direction it is moving in. Again, if a case of maintenance is found, considerable attention should be devoted to explaining why it happened.{6}

{6}Tendler, pp. 144-145.

4.3.4 Examining the Effect of Assistance Activity

This review of rapid impact evaluations has stressed their exploratory nature, their focus on a variety of impacts, and their consideration of alternatives in designing or modifying an activity. Although these characteristics imply a move away from testing simple, linear causal statements, it is still useful to examine the impact of a given assistance activity and how it contributed to the proposed causal chain. Although it is probably impossible to isolate the effects of any single development assistance, it is important to consider how it meshed with other elements in the environment. Drawing comparisons is the best way to do this. An irrigation evaluation, for example, could compare project areas before and after the irrigation projects were begun (before/after), or could compare them with similar areas that have no irrigation projects or different kinds of irrigation projects (with/without).

Before/after studies are appropriate when it is possible to get information about the "before." Unless adequate baseline data are available, this generally means retrospective interviews with local experts and beneficiaries. Some studies, however, have found existing baseline data that have been used to supplement reports from key informants.

> A study of rural roads was able to collect data about conditions before and after the roads were built. For example, the evaluation team found data on traffic counts and net farm income prior to the road improvements. Because the data were not sufficiently disaggreated, the team had to make assumptions about the data before using them.

With/without studies are appropriate when information can be collected about similar areas. It is often hard, however, to find other communities that are sufficiently similar on the important variables. The key words here are "important" and "sufficient." Which characteristics of a community are important in influencing the results? If size is presumably a factor, then a with/without study should compare communities of similar size to determine the impacts of a project. If size is thought to be relatively unimportant, however, then one would not need to find communities of the same size but would focus on other, more relevant characteristics. Similarly, a team will have to make judgments about how similar the communities need to be. Marginal differences may or may not be critical. Some studies avoid the issue by not explaining the basis of their comparisons. For example, one study of an irrigation project interviewed project beneficiaries and residents in another community about their satisfaction with life. Since the study does not state on what basis the control group (the residents of the other community) was selected, it is hard to evaluate its conclusions. Sometimes useful comparisons can be made within a community, for example, between those receiving and those not receiving project services.

Despite these evident difficulties, some impact studies have been very imaginative in drawing comparisons, as the following passage illustrates:

> An evaluation was designed to examine the results of a large project to reclaim land for agriculture in three different communities. In spite of the fact that 20 years had elapsed since the end of project funding, the study succeeded in drawing several comparisons.
>
> 1. In reviewing early documents on the projects, the evaluation team found pictures of the communities during the first years of the project. The team used these to constitute the "before" and then took pictures of the same places today. The team found, for example, that in one of the three communities, the early pictures showed that all the homes were one story high, whereas in the present community the buildings were mainly three stories high. They interpreted this change as an indicator of the community's efforts to improve itself after the project funds were withdrawn.
>
> 2. The evaluation team hired two local social scientists to interview a sample of families in the three communities. By interviewing both parents and children the team was able to compare the attitudes of those who remembered the early days of the project and those who did not.
>
> 3. The team found that the three communities had fared very differently and thus was able to compare the community where the project had a real and positive impact with the other two where few lasting results were found.
>
> 4. The team found maps of the original projects that specified the land use at the time of the project. In conducting the field study, team members reconstructed maps showing land use 20 years later and were able to document the amount of land preempted by the government and removed from agriculture.
>
> 5. The team learned of an existing study of the productivity of similar communities where land reclamation had not been tried. By including this study in their study design, team members were able to compare their findings with this study to determine whether land reclamation had made a difference.

A variation on with/without designs is to examine how an intervention works in different contexts. This is useful when it appears that an intervention is more successful in some settings than in others, and the point is to find out under what circumstances it has the greatest impact. Often

these circumstances will comprise many factors. In this case the comparisons may be useful to distinguish among the effects of different sets of variables, even if a specific element cannot be identified. Thus, a study may conclude that a project was more effective in villages closer to urban areas than in isolated villages. This may be operationally useful in designing future projects even if the study cannot isolate which factors were most crucial.

> A team studying a potable water project speculated that maintenance might be more efficient in a more closely knit community. To pursue this working hypothesis, team members designed the study to compare projects in high mountain areas, where the communities were more self-contained, with projects in more developed areas. They found, as proposed, that more maintenance was carried out in the former communities.

Another important consideration is how the project itself varies in different settings. The same project, for example, may use different technologies, different delivery systems, or different kinds of staff in different villages. By comparing the impacts of these project variants, we may well gain important knowledge about improvements in future project design.

4.4 Selecting Techniques and Resources

4.4.1 Existing Documents

Original project documents, particularly the logical framework, can be useful resources for conceptualizing impacts and causal factors. The logical framework indicates the original assumptions and hypotheses behind a project's design. Even if these have proved to be unrealistic, it can be helpful to refer to them and note in what ways they did not stand up. Materials documenting the initial steps in the process are also an important resource as noted earlier. Discussion papers and workshops bring together a variety of perspectives and experiences and often develop specific conceptual frameworks.

4.4.2 Alternative Perspectives

We know from learning theory that people get locked into perspectives and that it is often difficult to approach a project with fresh eyes. In addition, team members will usually be predisposed to look at a project from the perspective of their particular discipline. There are two ways teams can overcome this tendency -- one concerns how the team relates to the development community, and the second concerns team activities that stimulate members to look at problems in new ways.

4.4.3 Relationships With the Community

One way to look at a problem differently is to make contact with various groups within the project context, particularly those with presumably different perceptions and values. The concept of "stakeholder" is useful here -- it suggests identifying any group with a "stake" in a project and identifying its views. These views can stimulate consideration of different perspectives.

> One team studying rural health centers interviewed local government officials, health care providers both in and outside of the centers, consumers, and insurance cooperative officials. Team members also visited two health care centers not funded by the project to gain a better understanding of the problems in running rural health programs.

4.4.4 Team Efforts

Brainstorming. Topic coordinators or team leaders can structure the team's working sessions to avoid "groupthink" or premature consensus before all the alternatives have been explored. Various brainstorming techniques can be used to encourage each member to contribute to the discussion, rather than allowing one or two to dominate. The "nominal group process" technique, for example, requires each team member to write down ideas about a specific topic. Members of the group then take turns presenting their ideas, which are written on newsprint. Only when everyone's ideas are on view does any discussion take place, at which time the group tries to cluster and prioritize the different points. This may be a particularly useful technique if some of the members have extensive experience in the type of project being evaluated, because they might otherwise inhibit others from raising different perspectives and asking useful questions.

Being Provocative. One team leader wanted team members to look at a topic from a fresh angle. The theme of the workshop was on stimulating the private sector. He began by saying, "Let's imagine what a Burmese (Buddhist) would do if we gave him some money. From his perspective it would make sense to either build a pagoda, or go into smuggling. Given the local economy these are his only two viable economic strategies. What does this tell us about ways to encourage private sector initiatives?"

Developing a Matrix. Consider two different matrices, a checklist and a conceptual matrix.{7}

A checklist matrix, as its name implies, is used to collect similar information about several different groups or sites. For example, consider a study of a health care project in which project staff, users, and health care professionals were all interviewed.

Attitudes About	Users	Project Staff	Health Care Professionals in Area
Services Offered			
Accessibility			
Supplies			
Education			

Instructions: For each cell, indicate whether the respondent feels the item is inadequate, adequate, or ideal.

Whereas a checklist matrix primarily serves to structure the research and report data, a conceptual matrix can be more interesting by suggesting different relationships. Select two dimensions of a problem, each of which can assume varying values. Then place them in a matrix and speculate about the dynamics in each of the cells. What form will the interaction take? Which cells are more likely to occur? A very simple matrix could relate different degrees of irrigation project maintenance with a person's position.

		Location of Farmer: Head of System	Location of Farmer: Tail End of System
Participation in Maintaining the System	High		
	Medium		
	Low		

The point of the matrix is to get people to think about why those at the head or foot of an irrigation system might be more or less willing to help maintain it.

Identifying Actual Treatment. It is easy to take a project's description at face value rather than examine what is actually being done. For example, studies of several health clinics cannot assume they are all doing the same thing. One might offer free services to children in a family, another might stress education in preventive health care, and another might focus on family planning. Thus, it is important to specify the activity that is being examined and not rely on labels of activities or projects.

Developing Categories. One of the most useful conceptual techniques is to think about ways to categorize the most important variables. For example, studies that try to incorporate contextual factors and external variables often find it difficult to set limits on what they are examining. One technique is to consider how to categorize different environmental factors. One could distinguish among environmental factors that provided positive resources, those that were constraining, and those that were neutral. Such a strategy would

provide a way to organize variables and select which to examine.

Tendler, writing about evaluations of participation in projects, notes that participation is usually defined very simplistically and is almost always assumed to have positive impacts. She suggests distinguishing among different forms of participation in project activities as follows: (1) representative participation, where those involved try to represent the different interests in the community; (2) top-down but sensitive participation, where participation is by elites who are responsive to the community; and (3) participation by elites who continue to act in their own interests. The effects of participation will vary considerably depending on the type. The value of this conceptualization is that a study can consider the impact of a particular kind of participation, rather than assume that participation is good in and of itself, or that participation by elites is always dysfunctional.

{7}Matthew B. Miles and A. Michael Huberman, Qualitative_Data_ Analysis (Beverly Hills, California: Sage, 1984), pp. 83-111.

4.5 Selecting Measures or Proxies for Variables

The search for appropriate measures is often the most challenging and difficult part of a study. There are two problems in selecting measures. The first is finding measures relevant to needs to be learned, and the second is finding ones for which data can be gathered within a short timeframe. For example, almost every development assistance activity tries to improve local standards of living. What needs to be determined are what indicators will measure standards of living and what data are available?

4.5.1 Selection Criteria

Simple and Feasible. Measures are often more complex than they need to be. Often the impacts are so obvious that complicated measures of impact are unnecessary. For example, a study of an irrigation project easily determined by observation whether irrigation canals were delivering water. Another was able to determine if fertilizer was delivered by looking at inventories and to determine if people were using available credit by finding out how many loans had been made. Robert Chambers refers to "optimal ignorance" and "appropriate imprecision."{8} These terms suggest that we should consider what we need to know, what we do not need to know, and how precise we need to be. It is important to avoid studies that are more complicated than necessary or the collection of data that are too complex to be readily analyzed. For these reasons Chambers recommends selecting simple, cost-effective proxy measures and

designing ways to do "rapid reconnaissance."

The challenge is to find measures for which data collection is feasible, but to be sure to collect information from all groups in the population. In selecting cases, the simplest strategy, may not be the best. For example, some measures are subject to such problems as "dry season" and "tarmac" biases. Researchers or observers will often visit places during the dry season or only go where there are paved roads. The problem is that they fail to collect information from those who may have the greatest needs, and thus they end up with unacceptably biased measures. Thus "simple and feasible" apply to the quality of the measure and not to the accessibility of the subjects. (Sampling will be dealt with in Section 5.)

Proxy measures are less directly connected to an impact or response and are often chosen precisely because they are simple and feasible. Increases in income and attitudes of well-being are fairly direct measures of higher living standards, but like many such measures may be difficult to determine. Proxy measures are less direct, but are more visible, usually quantifiable, and more readily available. The question is, when are proxy measures appropriate, or when are they unacceptably "quick and dirty," to use Chambers' terms? The answer lies in balancing the simple and feasible criterion with the other criteria listed below.

Validity. Measures have face validity if they are relevant to the questions being asked. Knowing that people attend a health clinic will tell you about usage, but it may not tell you how much people like the clinic if they have no alternative. If the research is trying to determine attitudes, data about usage would rate low on face validity.

Reliability. To what extent will the results be the same, no matter who collects the data or when the measure is repeated? In other words, is the measure free of the data collector's bias? Would another team collecting the same information get the same answers?

Appropriate to Context. Do the measures depend on appropriate assumptions? George Honadle notes that often this is not the case.{9} For example, the number of tin roofs in a village may be used as a proxy measure of prosperity. This may not be a valid measure in a community where members have improved their homes to impress a visiting dignitary or in a place such as Kenya, where herders place a higher value on cattle than on acquiring tin roofs. His point is that teams need to clarify the assumptions behind their measures, rather than grasp too quickly at "simple and feasible" data.

Precision. Measures can vary in their precision. Interval measures are used when a uniform measurement scale, such as weight, temperature, or income, can be applied to a set of cases or observations. All values on an interval scale are equidistant; that is, the difference between 20 and 21 pounds is the same as the difference between 30 and 31 pounds. Interval

measures can be very precise and are usually desirable when such data are easily available.

Given the constraints of doing evaluations in the development context, interval data are often unavailable, difficult to collect, or unreliable. The alternatives are to use ordinal or nominal measures. Ordinal measures indicate that one group of cases has more of a characteristic than another group, allowing them to be ranked. For example, if actual income figures cannot be obtained, proxy measures of income can be used, such as observations of housing conditions. The homes in a community can be observed and judgments made about whether their inhabitants fall into a low, medium, or high income group. Nominal measures essentially categorize the data into groups, such as men and women, tenants and landowners, and so forth. They classify cases rather than measure them and hence are the least precise.

One technique that is useful when precise interval level data are unavailable, or when relatively complex judgments are called for, is to set up a simple ordinal scale with three (or more) values. Then key informants or simple observations are used to identify the value for each observation. For example, to determine if a project had positive social benefits, one could ask people to assign a value to them: three if the benefits clearly outweighed any costs, two if the net benefits were marginal, and one if the costs clearly outweighed the benefits. If the observations do not warrant making three distinctions, the team may decide to create a nominal dichotomy, coding observations by whether or not they benefit the poor.

> A study of rural roads was designed to determine if the roads benefit the poor. The authors devised the following scale. Roads that provide access to isolated mountain villages received a three (the poor are primary beneficiaries); roads that upgraded dirt roads and serve towns were scored as two (others besides the poor benefit); roads that serve large plantations were scored one (the poor do not benefit directly). The measures were treated as ordinal measures of the "extent to which the project benefited the poor."
>
> One study of health projects used records from the Ministry of Public Health (MOPH) to determine how potable the local water was because they had no time to test the water themselves. MOPH data were reported in nominal categories, "clean" and "unclean," and thus the team had to analyze the data using less precise techniques than they had wished.

{8}Robert Chambers, "Short-Cut Methods in Social Information Gathering for Rural Development Projects," in Putting People First, edited by Michael Cernea (Washington, D.C.: World Bank, 1983).

{9}George Honadle, "Rapid Reconnaissance for Development

Administration," World Development 10 (1982): 633-649.

4.5.2 Examples of Measures Used in Impact Evaluations

A team measured increased income in a community by the increases in payments for school tuition.

A team studying the effects of pesticides on public health went to local hospitals to see if there were any records of pesticide poisoning. When team members found no such records, they concluded that it was possible that medical personnel did not know how to diagnose such poisoning. Thus, they began to look for changes in the incidence of other problems such as spontaneous abortions or eye diseases. They could not use these as definitive measures of the effects of pesticides, but concluded that they indicated problems that should be looked at more closely in the future.

A team studying agriculture research centers was asked to measure how sustainable the centers were. Team members felt that sustainability meant not only the survival of a project intervention, but also an ability to adapt to changing circumstances. Thus as one measure of sustainability, team members asked research directors how they had adjusted research agendas given information about farmer needs.

4.5.3 Two Guidelines in Selecting Measures

Tradeoffs. Tradeoffs will almost always have to be made among the criteria for defining measures. For example, in-depth interviews may indicate how beneficiaries really feel about a project and hence may be high on validity. However, such interviews may be lower on reliability, because interviewers may influence what answers are given. Counting tin roofs as a measure of income may be both feasible and simple but may not be as valid as other measures.

Triangulation. Wherever possible, two or three measures or sources of data should be used. Ideally the measures will draw from different kinds of data. For example, observations of an irrigation system and interviews of beneficiaries could be used, rather than two measures both based on observations. This is more likely to provide data that are strong on different criteria.

4.5.4 Sample Questions for Describing Impacts

The following sets of questions have been proposed by those administering the impact evaluations.

Beneficiaries

-- Who were the intended beneficiaries?

-- Who actually benefited?

-- Were benefits equitably distributed?

-- Did both women and men benefit?

Social and Economic Impact

-- What social and economic changes occurred?

-- Were there any changes in organizations? New organizations?

-- Were there changes in the number or use of private enterprises?

-- Were there changes in productivity, income, employment?

-- Was the host country's macroeconomic situation affected?

-- Was there a change in rural-urban migration patterns?

-- Were there changes in the use of social services?

Were women and men involved or affected by these changes differentially?

Political Impact

-- What, if any, was the impact on political institutions?

-- What, if any, was the impact on political processes or policies?

-- Was there a change in management capability?

-- Was there an impact on the host government's budget?

Technological Impact

-- Was there an impact on indigenous technology?

-- Did innovations spread beyond original sites?

-- Were resources adapted to local conditions?

-- Was there an increased dependency on external resources?

Sustainability

-- Did intended impacts appear to be long-lasting, durable?

-- Did impacts spread to other areas, groups, institutions?

-- Did fees for project goods or services cover their cost?

-- Were activities continued and maintained?

4.5.5 Sample Questions for Identifying Causes of Impact

Role of Host Government

-- How did its policies, procedures, and priorities affect the activity?

-- How did its institutional capacities and practices affect it?

-- What was the effect of the overall economic situation? Were the recurrent costs, foreign exchange requirements, or external costs beyond (within) the host government's financial and economic abilities to sustain? Were the project costs more (less) than planned?

-- Was there a consistency between A.I.D. and host government policies, and how did this affect the results?

Beneficiaries

-- Did the extent (lack) of participation by target beneficiaries affect the impact?

-- How did the historical, social, ethnic, or cultural characteristics of beneficiaries affect the results? Was the assistance consistent with the prevailing social and cultural mores of the country?

Design

-- Was the design appropriate?

-- Were resources provided appropriate to the level of technological development of the host country?

-- Were there complementary inputs and services and did these affect the results?

-- Did the timing and sequence of inputs affect the results?

Implementation

-- How did A.I.D. and contractor performance affect the impact?

-- Were the activities managed well (poorly)?

-- Was the quality of the technical assistance adequate?

-- Did the implementors try to accommodate host government values and operational styles?

5. CONDUCTING EVALUATIONS: DATA COLLECTION AND ANALYSIS

This section examines several aspects of collecting data, including sampling and data collection techniques, with an emphasis on interviewing. It also considers the equally important issue of analyzing the collected data.

5.1 Selection of Site Visits

When teams arrive in a country (sometimes before the team arrives in a country), one of their first tasks is to decide on the specific project sites or activities relevant to their study. Because they are operating under severe time limitations, they have to be selective in choosing sites to visit. The criteria they use for selection can make a major difference in the validity of their findings.

Bias can easily creep into the selection of site visits. Logistics may make it easier to visit some sites than others. Missions may steer researchers to those sites they feel best about. One evaluator reflected that the major problem he faced was the "Potemkin village problem," referring to the classic instance in which Russian bureaucrats constructed facades of villages for visiting dignitaries and then dismantled them when they had left.

More subtle biases also can affect selection. Interviews in a village may take place at a time when most of the males have migrated to cities for work. As noted earlier, Chambers writes compellingly about the common biases that can afflict data collection such as a "tarmac bias" and a "dry season bias." Because of the short time for the evaluation, even conscientious researchers are more apt to interview or observe those who live near paved roads than those in remote areas, and more apt to approach people during the dry rather than the wet season. In these cases, researchers will get a more positive view of rural conditions than they would if they visited people in more remote areas during the wet season when poor health may be more evident.

If a sample of sites can be randomly drawn from the project, and if it is of sufficient size, then one can be confident that the characteristics of the sample reflect the characteristics of the population with certain known probabilities of error.

A study of water projects used a sample of 31 sites. The sites were stratified according to whether their water came from a well or spring, and by district. Within each of four

districts a number of well and spring projects were randomly selected, proportional to the total number of each kind of project in the entire assistance package.

However, random sampling is often impossible and may not even be desirable:

A team studying potable water projects in rural villages visited 30 sites selected on the basis of a random sample. This meant that team members had to visit three sites each day. Although they got very reliable information, it was a very tiring and time-consuming activity, and meant that the team could spend very little time in each place and had to ignore other issues.

Generically, sampling can be considered as an effort to select sites or activities that appropriately represent characteristics of the total group that are relevant to the evaluation. In a few cases it may be possible and appropriate to select the sample randomly. In other cases the sample can be drawn from different groups in terms of other criteria of variation in a community. The key is to define the critical dimensions that a team thinks may be relevant to describing or explaining a particular impact and purposively to select units illustrating these dimensions. The strategy is to compare the impact of the project among these different groups rather than to generalize across a population.

For example, an evaluation of irrigation projects might interview both those at the head and those at the foot of an irrigation system. The point would be to compare how the system affected these two groups. Or if the study is examining the utilization of a health clinic, it might compare groups with different access to the clinic. Similarly, cases can be chosen to explore the relevance of different contexts. If land tenure is presumed to be an important factor in explaining maintenance of irrigation systems, three communities could be chosen representing different land tenure arrangements. Again the purpose is not to arrive at a statistical measure of the relationship between types of land tenure and maintenance activities, but to compare maintenance in three different situations.

In a study of health clinics a team felt that the region of the country might be an important factor in explaining their impacts. Team members went to three villages in each of three regions. They felt that they gained a sense of how clinics function in different kinds of areas, and that the relatively small number of cases gave them an opportunity to spend more time in each place and to explore issues in some depth.

It is clear from this discussion that conceptualizing the critical variables in a study is very important. It is even more important if one cannot randomly select the sites to visit.

One team looking at irrigation projects wanted to know what dimensions the local government believed were most important. When team members found that local officials categorized projects according to whether or not loans were repaid, they looked at several projects in which the repayment rates were high and compared them with ones in which they were low. When they analyzed their results, it appeared that climatic zone made a difference in explaining whether or not the loans were repaid. It was too late to select a group of projects according to different climatic zones, but they could suggest that this dimension should be explored in a further study.

Even if the presumed differences do not turn out as expected, it is still possible to learn something.

One team was asked to do an impact evaluation of 50 irrigation projects. In order to pick a sample, team members began by speculating about the critical factors in irrigation. They used size and source of energy and selected a group of irrigation projects within each group. But they found that neither of these factors had made any difference in the results. This apparently negative finding actually told them a great deal.

5.2 Data Collection

The issues associated with selecting appropriate measures which were discussed in Section 4.5, are also important in deciding how to collect the actual data to be used. Chambers suggests several strategies that he believes will increase the chance of getting appropriate data:

-- Taking sufficient time to observe what is going on

-- Being aware of biases, such as talking only to men

-- Appearing unimportant (More is likely to be learned by those who are not perceived as important officials or visitors.)

-- Emphasizing listening and learning (The simplest people will often have important and useful impressions about an activity.)

Essentially there are three ways to collect information:

-- Using already available data

-- Questioning people through interviews, surveys, or questionnaires

-- Observing through field visits

5.2.1 Available Data

Sometimes data are available from the host government, from the project implementing agency, from earlier studies done by the Mission, or even from private consulting studies. Such data will have been collected for purposes that may or may not be consistent with the evaluation, and may be biased or contain major errors.

> An evaluation of PL 480 funds obtained considerable data from the government relevant to evaluating a health project. But when the team looked at the data carefully, it found a high error rate on relatively simple variables such as weight and age.

At the same time, available data are often an important basis for longitudinal and before/after comparisons. It is almost always worth asking people what kinds of data exist and what they routinely collect. Aerial surveys can quickly show land use and irrigation patterns and can open up relatively inaccessible areas. Health clinics may keep records that provide useful information.

> One team visited a health project and examined records covering a number of years. Team members noted a high number of children from the same families. They compared the records on the older children in a family who had come when the clinic first opened with the records on younger children from the same family who entered the clinic at a later time. This gave them a rough measure of the effects of the clinic in teaching families good health practices.

Cost Data. Although often complicated, and sometimes disorganized, cost data are an essential basis for cost-effectiveness, cost-benefit, and other economic analyses. Although such analysis should not be the sole criterion for project evaluation, it has an important bearing on project design and program strategy decisions. When extensive economic analysis is envisioned, appropriate expertise should be included on the team, as well as a pre-fieldwork effort to assess or collect appropriate cost data.

> A team examining health clinics found that another study on similar facilities had been done already. This meant that the government had already set up an accounting system that could give the evaluation team the kind of cost information it wanted.
>
> Another team knew ahead of time exactly the kind of cost data it needed, and contracted with the Mission to get a local researcher to collect data on the cost of certain commodities. Even with this head start, team members found

that the data were not always clear and that they had to do a lot of rechecking.

CDIE is currently exploring improved methods for incorporating economic analysis in rapid impact studies.

Income Effects. It is usually difficult to determine the income effects of a program by asking people about changes in their financial situation. If the information is important, other sources of this information should be sought. For example, are tax records available? Does the Ministry of Agriculture keep records on output or on marketable surpluses? Or someone from the village can be hired to shop in various villages to check and compare prices.

5.2.2 Questioning People

Extensive sample surveys are seldom possible in the limited time allotted to impact evaluations. Therefore, evaluators tend to rely heavily on in-depth interviews. Most interviews will be conducted among key informants in the community, those associated with the project, or with beneficiaries. Such interviews are valuable because they provide information on the reasons behind behavior, they are sensitive to different points of view, and they can probe contextual and process issues.

In-depth interviews generally require more forethought than they are given. They are a major example of what are often referred to as "intrusive data," meaning that it is very easy for the interviewer to influence or intrude on the answers. Thus, an obvious rule of thumb is to think of the interviews as a way to find out what is in someone's mind rather than to be suggestive.{10}

A team may try to use local people to assist in designing the questions because they can be helpful in sensitizing the team to local viewpoints. In some cultures it is considered offensive to be asked about educational background or the family's role in a business. Questions about income generate notoriously inaccurate results, because people are embarrassed by the question, they define income differently than others do, or they honestly cannot remember. Expatriate interviewers are frequently perceived as hostile or fault finding, and thus it is important to find ways to put people at ease and reassure them about the purpose of the interviews.

Because in-depth interviews presumably allow team members to get "close to what is really happening," it is easy to think of them as the most valid information one can gather. Such data have their own pitfalls, however. In effect they produce three different kinds of information: (1) descriptions of events, (2) interpretations of events by the subject who inevitably wants to appear favorably, and (3) interpretations by the interviewer of what is said. It often requires a good deal of skill and cross checking with other respondents to sort out these three

levels and obtain a clear description of what happened.{11}

Interviews can vary from being open-ended conversations -- "Tell me what you think about the project" -- to closely structured interviews that follow a list of specific questions. The tradeoffs between these are obvious. The less structure, the more likely one is to get a full understanding of how the respondent feels about a situation and the meaning attached to words. The more structure, the more the interviewer is apt to cover issues laid out ahead of time by the team and to gather comparable data that can be aggregated with the results from other interviews. Some degree of structure is almost always desirable with rapid impact studies, particularly when there are several interviewers. And yet it is important to depart from the protocol to probe and follow up on answers, to ask "Can you tell me more specifically what you mean?" or "How do you think that occurred?"

Patton{12} lists the following guidelines for interviews:

Avoid dichotomous "yes/no" questions. They seldom produce the desired information, and they make people feel they are being quizzed.

Questions should be single, otherwise they create confusion.

Use terms the respondents are used to. For example, they are probably not used to "PVO" for a private voluntary organization.

Avoid asking "why" questions; rephrase them to get at the kind of reason you are interested in.

If the interview is long, group the questions and introduce them. For example, "Now I am going to ask several questions about the use you made of the new well."

Probe. Ask for more details, ask them to elaborate, ask them to clarify a point.

Encourage them. For example, "These comments are very helpful." "Now we are halfway through the questions."

{10}Michael Patton, Qualitative Evaluation Methods (Beverly Hills, California: Sage, 1980).

{11}Jon Van Maanen, "The Fact of Fiction in Organizational Ethnography," Administrative Science Quarterly 24 (1979): 539-550.

{12}Patton, pp. 195-250.

5.2.3 Interview Strategies

Pretesting Questions. Ideally this should be done while the team is still together outside of the capital city, so they can discuss and make needed changes.

Selecting Key Informants. Often the most useful sources of information are local officials or leaders in the community. They will probably be biased because they are commonly better off than others, but they can be an important resource. An attempt should be made to talk with officials who hold different views and have different perspectives. When possible, interviews with mid-level officials should be conducted apart from higher ranking officials.

Group Interviews. These can be useful to get a range of opinions quickly. People may feel more secure in providing sensitive information in a group. For example, women may feel more comfortable discussing infant feeding practices in a group than as individuals.

> A team studying water projects interviewed beneficiaries, trying to reach both men and women. They went to each site and gathered the available beneficiaries into groups where possible, enabling them to reach a considerable number of people in a short time. Team members also tried to group the men and women separately to encourage the women to speak more freely.

Working With Translators. Even if local people are part of the evaluation team, additional translators can be useful. Otherwise, all information is funneled through a single source. In some cases local people tend to gloss over sensitive questions that an expatriate might feel free to ask. In one study, for example, local interviewers felt uncomfortable asking local officials about budget information, whereas the expatriate team members were able to do so.

Unless the translators are very skilled and knowledgeable about the kinds of information that are needed, it is usually desirable for the translator to report the answers as the interview proceeds. This allows a team member to suggest probes and follow up questions on the spot. If translators write up their notes and give them to a third party to translate into English, important information and impressions inevitably get lost. If the translators have conducted interviews on their own, it is better to interview them about the results rather than wait for them to be translated.

Group Meetings. Sometimes a group of officials can be brought together to consider a series of specific questions. This strategy not only produces data for the team, but has the added benefit of getting officials to share information among themselves.

> A team wanted to know which government agency was

responsible for specific steps in planning and implementing projects. Team members brought together top officials from the relevant agencies and presented them with a matrix, part of which is given below.

	Planners	Engineers	Contractor	A.I.D.
Project Selection				
Site Inspection				
Cost Estimates				

Participants filled in each cell according to who had major responsibility, who was consulted, and who was informed.

5.2.4 Observations

Sometimes the simplest, most effective data are collected by observing activities or projects first-hand. Visits to irrigation projects can provide information on how well they are maintained. Walks along canals can provide opportunities to talk informally with farmers. Often it will be helpful for the team to develop a data collection form, so that all members will look for similar information and will classify observations in comparable categories.

A study of water projects was based primarily on observations of the projects. The team developed a data collection form to gather comparable information. It contained the following headings: water source, physical characteristics of site, description of settlement pattern (discrete village, dispersed population,...), description of beneficiaries, whether water source is working, description of improved water source (number of users, accessibility, water quality...).... Once all of the sheets were filled out it was relatively easy to aggregate the results and look for patterns.

5.3 Analysis

One of the easiest pitfalls in any field research is to get consumed by the field, to get caught up in collecting more and more data, to become what some might call "field happy." Analyzing the data can all too easily take a back seat and be done in an offhand and hurried way. Several steps can be taken to encourage more attention to analysis.

5.3.1 Scheduling Time

Team members have found that they need to plan their schedules so they will have time to meet periodically to report on and discuss their findings. A member of one team reported that group discussions allowed team members to

crystallize their impressions and be more directed in what they looked for.

A team studying potable water projects met to draw up a list of questions for beneficiaries. Team members traveled together to a community, fanned out to conduct their interviews, and then met in the evening to reflect on the information they were gathering. The next day they traveled as a team to a second region and followed the same pattern. The group discussions became a critical part of the research and led them to sharpen their focus and impressions. Essentially, they found that they were writing the report as they proceeded.

5.3.2 Summary Devices

Team members might develop a matrix or chart listing the different dimensions they are looking at, and fill this in as they proceed. This encourages them to look for patterns.

In studying a potable water project the team developed sets of physical and social indicators to measure how well the projects were working. Physical indicators included (a) adequacy of source of water, (b) whether incline of the pipe was adequate for the pressure, (c) puddles near faucets indicating leakage. Social indicators included (a) hours people spent fetching water prior to the project, and (b) how they used the time saved after the project. The team developed codes for different values for each indicator. For example, adequacy of source of water was given a 1 for inadequate or a 2 for adequate. Then the team constructed the following matrix, which team members filled in as they went along.

	Physical_Characteristics				Social_Characteristics		
SITES	_A_	_B_	_C_	_D_	_A_	_B_	_C_
.....							
.....							
.....							

Each evening when team members met they shared their information, compared patterns, and speculated about the reasons for the results. Because they had collected the same kinds of information, they were able to be much more precise in their comparisons and, hence, in their analyses.

5.3.3 Displays

Displays such as matrices are useful components of analysis. According to one study, field studies usually rely on narratives to "display" their results. However, used alone, narratives are both "weak and cumbersome." "Valid analysis requires, and is driven by, displays that are as systematically arranged as the questions at hand demand." Displays of quantitative data are common, but the "qualitative analyst has to hand-craft all such data displays."{13} Matrices and checklists are examples of ways to display qualitative data that make them readily visible and enable team members and readers to find patterns and relationships.

_____

{13}Miles and Huberman, p. 79.

6. COMMUNICATING RESULTS

This section will deal both with preparing written reports and, in less detail, with debriefing sessions for Mission staff and host governments.

6.1 Written Reports

Because impact evaluations are designed to inform those officials in A.I.D. who set policy, assign resources, and design activities, the results need to be readily accessible to busy officials and easily grasped. In addition, results should be written to facilitate comparisons among activities. For both of these reasons, CDIE has developed the following criteria:

They should be engagingly written (avoid being too colloquial or impersonal and ponderous).

They should be concise, with 15 pages for the summary report and not more than 100 pages overall.

They should be carefully organized.

Analysis must be supported by evidence.

They should follow the outline and guidelines described below and in CDIE's Publications_Style_Guide.

6.1.1 Report Outline

Title Page

Table of Contents
Foreword
Acknowledgments (optional)
Summary
Preface (optional)
Project Data Sheet
Glossary
Map
Body of Report
1. Setting
2. Description of Activity
3. Impact: Findings and Analysis
4. Lessons Learned and Policy Implications

Appendixes

Bibliography (optional)

6.1.2 Elements of the Report

With the exception of the Foreword, which is written by CDIE staff, all sections of the impact evaluation are the responsibility of the team.

Title Page. Include the title (with location and activity); members of the evaluation team and their affiliations; U.S. Agency for International Development; the date; and the disclaimer that "The views and interpretations expressed in this report are those of the authors and should not be attributed to the Agency for International Development."

Summary. This should be no more than two single-spaced pages. It should be capable of "standing alone" without reference to the rest of the report.

Preface. This includes introductory comments from the team and acknowledgments. If acknowledgments are lengthy, they should be placed in a separate Acknowledgment section following the Foreword.

Data_Sheet. This should list pertinent data such as the title, project and loan numbers, amounts, terms, dates, and purpose.

Glossary. This should identify and define any terms particular to the location, culture, or activity. Currency equivalents and weights and measures also may be included.

Body of the Report. These sections (1-4) constitute the basic narrative of the report and should be written while the team is in the field. The body should not exceed l5 single-spaced pages. Supportive data that cannot be accommodated within this limit can be attached as Appendixes, but the overall length should not exceed 100 pages. The analysis should be

carefully reasoned and based on clearly presented evidence.

1. Setting. This section should be limited to one or two pages. It is always tempting to write more, but that leaves little room for the findings. This section should briefly summarize the initial conditions against which the team will assess any changes that occurred during and after the activity. It should contain a description of the relevant historical, political, social, economic, institutional, and/or policy issues that produced the need for the activity or affected its impact. Quantitative data should be used if available.

2. Description of the Activity. The project or program should be described briefly, including its goals, purposes, and (if relevant) its assumptions. A brief history of the implementation experience should be provided.

3. Impact: Findings and Analysis. This is the major part of the report and should be approximately 10 pages in length. It should identify impacts or changes that occurred, whether these could be attributed to the assistance activity, and the other variables that seem relevant. Supporting data should be provided.

4. Lessons Learned and Policy Implications. This is a critically important section and also one of the most difficult to write. The point is to reflect on the implications for A.I.D. policies and programming and for replicating activities in other contexts. This section is most useful when it is able to address policy and design problems that the Agency is currently considering. Were there any lessons related to the conceptual framework or the common questions that the teams developed at the workshop? In practice, some reports list a series of lessons, whereas others select two or three lessons and discuss them in a paragraph or so. Lessons learned about design and management of activities should be included, as well as policy lessons.

It is easy to draw obvious or trivial conclusions, such as "Training should be adapted to needs." Analytic statements are usually more helpful, such as "It is difficult to adapt training to local needs because that greatly adds to the cost of the training, and many host institutions do not feel that such adaptation is particularly important." Or, "Training can best be adapted to local needs if working units receive training together, and if the training is designed around some specific tasks in which they are involved." These elaborations are more specific and give the reader some insight into what to do with the lesson.

What may appear trivial to some is not necessarily so to others, and what may appear obvious is not necessarily heeded. For example, a report might offer the lesson that maintenance of irrigation projects is very important. From one perspective this is a very obvious point, and has been said before. At the same time, designers and implementors still neglect it, and it is useful to document its importance. Even so, the lesson could be made more useful by stating why it was neglected or practiced in

this case or what was learned about providing incentives to encourage maintenance. "Maintenance of water projects is unlikely unless there is a mechanism for charging users a tariff and ensuring that it is collected."

Appendixes. One Appendix should describe the team's methodology. Others should be included insofar as they contribute directly to the substance of the report. In a few cases appendixes might be used to report opinions of dissenting members of the team or rejoinders by USA.I.D. or host governments. Appendixes can include further supporting data, background studies on aspects of the study, photographs, notes on the team members, or lists of people contacted.

6.1.3 Logistics of Report Writing

Teams should consider the possibility of using the word processing facilities at the Mission to enter their report directly onto a diskette that they can take back to their home office. If they spend their last few days in the capital city and are willing to use the facilities at off hours, this may be possible.

The team leader is responsible for the report as a whole but can assign different sections to individual team members.

Reports are reviewed for the cogency of their argument and presentation and are edited for clarity and consistency. The accuracy of the findings, however, remains the responsibility of the authors.

Because of the difficulty of getting material typed in A.I.D./Washington, topic coordinators or team members should consider including in the scope of work of one of the contractor team members the additional task of typing one or two drafts of the report.

6.1.4 Style

Avoid jargon. Try to write in the active voice; it is usually more informative and forceful.

Consult the CDIE Publications_Style_Guide for specific guidance on format, grammar, and style.

6.2 Debriefing Officials

It is usually desirable to arrange a debriefing session prior to leaving the field. This will have to be preliminary because the report will not be final nor will it have been reviewed by

the Agency. Still, it is an occasion to provide the Mission and host government an opportunity to comment on the report. Its tenor would be, "These are some of the things we have found. Do they fit with your perceptions? Are you considering any changes we should know about? Are you struck by any other results we may have missed?"

One study of urban service projects organized a debriefing session around three broad headings: accomplishments, continuing problems, recommendations. The recommendations were put in two groups: specific, e.g., schedule payments to fit into stages of construction; general, e.g., reassess whether there are channels for coordinating activities. Mission staff and host government officials were given an opportunity to respond to both the observations and the recommendations.

To the extent that the research was guided by Agency priorities and formulated within a conceptual framework, these elements can be presented to the Mission together with any tentative findings. Those in the field may have comments based on their other activities and would gain by knowing what issues are being emphasized by the Agency.

6.3 Concluding the Evaluation Process

Team members return and complete their draft report. This is circulated within CDIE and to other interested parties. Team members may be asked to participate in debriefings. A synthesis report of all related field evaluations is eventually prepared and a conference is held. Then a final report that summarizes the individual reports and the conference results is prepared. As indicated earlier, these are published in a variety of forms to increase their visibility and utility to various members in the A.I.D. and development communities. Dissemination might include articles in Frontlines and Horizon (A.I.D. publications), in Development Experience Abstracts (a CDIE publication), as well as formal reports.

6.4 Responsibilities of Topic Coordinators, Team Leaders, and Team Members

6.4.1 CDIE Topic Coordinators

Develop background discussion and issue papers.

Identify project or program to be evaluated.

Schedule planned evaluation.

Send initial cable to Missions requesting approval for scheduled evaluations.

Select team leaders.

Assist team leaders in selecting team members.

Get CDIE Office Director approval for team composition.

Ensure that all team members are given relevant background documents to review.

Be thoroughly familiar with substance of background documents.

Consider how much structure is appropriate for the separate evaluations.

Organize and chair briefing workshop.

In cooperation with team leaders, send draft scope of work to Missions.

Respond to any communication from the field regarding team evaluation.

Plan pre-evaluation workshops and meetings, final conferences and seminars, and oversee publication of conference proceedings.

Arrange for any briefings to be held with A.I.D./Washington staff and with senior management.

Set and monitor deadlines to ensure timely publication.

Circulate draft report to members of the Working Group (if there is one) and other appropriate officials in A.I.D.. Collect comments and forward them to team leader.

Ensure that subsequent drafts deal with these comments.

Arrange for circulation of final draft to senior staff. When accepted, ensure that any final comments are incorporated in the camera-ready copy and arrange for publication.

6.4.2 Team Leaders

Direct impact evaluation, coordinating activities with topic coordinator.

Identify disciplines and expertise needed for team.

Identify resource people to participate in briefing workshop.

Select team members, in cooperation with topic coordinator.

Set travel dates in consultation with team members, field posts, and sector coordinator.

Help topic coordinator draft scope of work cable to Mission.

Facilitate team activities, clarify roles of each member, and arrange for team meetings.

Make necessary contacts with Mission.

Manage evaluation activities in the field.

Prepare a complete first draft of report prior to return to the United States.

Conduct briefing session with Mission before return, if one is planned.

Submit draft report to topic coordinator within a week of return to the United States.

Receive criticisms of initial draft and ensure that the final draft accommodates them.

Get final report to topic coordinator within 1 month after comments on first draft are received.

6.4.3 All Team Members

Review background documents.

Participate in briefing workshop.

Prepare individual scopes of work and participate in setting evaluation strategy during workshop.

Begin any data collection if that is warranted.

Ensure that travel authorization is prepared; ensure that passport is in order; get shots and visas; and make all personal arrangements for timely departure.

Conduct field evaluation as member of the team.

Write ssections of report and appendixes, as requested by team leader, in accordance with outline in this review and CDIE's Publications Style Guide.

Participate with team leader in briefing Mission and host government on findings.

Brief A.I.D./Washington offices and senior management, as arranged by topic coordinator or team leader.

Attend final conference.

Attend subsequent briefing workshops for new teams.

Rewrite sections of report as requested by, and in consultation with, team leader.

BIBLIOGRAPHY

American Council For Voluntary Agencies for Foreign Service (ACVAFS). Evaluation Sourcebook. New York, New York: ACVAFS, 1983.

ACVAFS. Approaches to Evaluation: Report of a Workshop on Impact Evaluation. New York, New York: ACVAFS, 1981.

Britan, Gerald. "Contextual Evaluation: An Ethnographic Approach to Program Assessment." In Methodological Advances in Evaluation Research, edited by R. Connor. Beverly Hills, California: Sage, 1981.

Bryant, Coralie, Louise White, Elisabeth Shields, and Therese Borden. Research in Development Management. Washington D.C.: Bureau for Science and Technology, U.S. Agency for International Development, 1983.

Chambers, Robert. "Short Cut Methods in Social Information Gathering for Rural Development Projects." In Putting People First, edited by Michael Cernea. Washington, D.C.: World Bank, 1983.

Honadle, George. "Rapid Reconnaissance for Development Administration." World Development 10 (1982):633-649.

Miles, Matthew B., and A. Michael Huberman. Qualitative Data Analysis. Beverly Hills, California: Sage, 1984.

Miller, Trudi. Public Sector Performance. Baltimore, Maryland: Johns Hopkins University Press, 1984.

Ostrom, Elinor. Strategies of Political Inquiry. Beverly Hills, California: Sage, 1982.

Patton, Michael. Qualitative Evaluation Methods. Beverly Hills, California: Sage, 1980.

Paul, Samuel. Managing Development Programs. Boulder, Colorado: Westview Press, 1982.

Tendler, Judith. Turning Private Voluntary Organizations Into Development Agencies: Questions for Evaluation. Program Evaluation Discussion Report No. 12. Washington D.C.:

Agency for International Development, 1982.

Van Maanen, Jon. "The Fact of Fiction in Organizational Ethnography." Administrative Science Quarterly 24 (1979): 539-550.